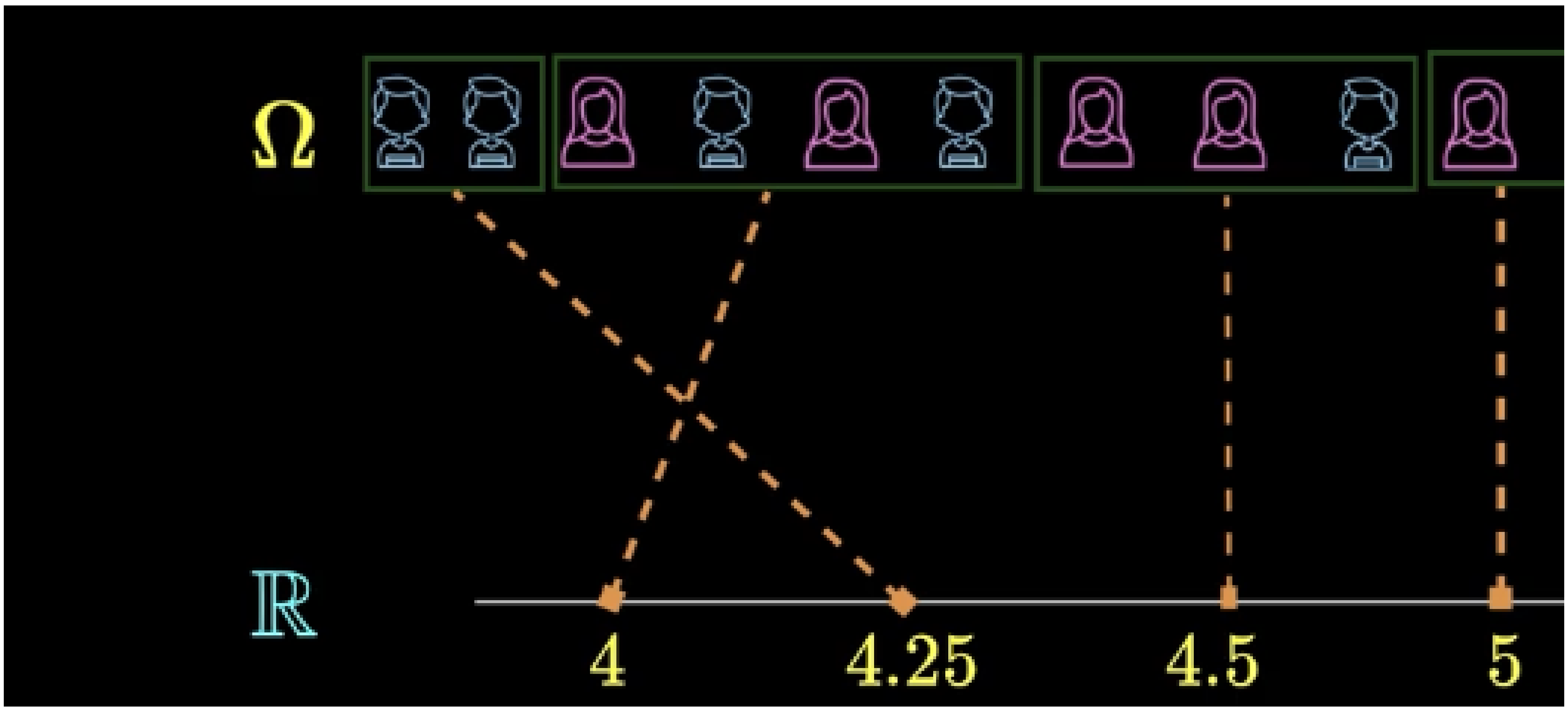


Week 16 : Random Variables

pending tasks	
type	

This module focuses on the numerical quantities associated with the outcomes of experiments.

Random Variable

def random variable - (t = 13:20)

A sample space (domain) can be mapped to different numerical quantities to answer questions of interest.

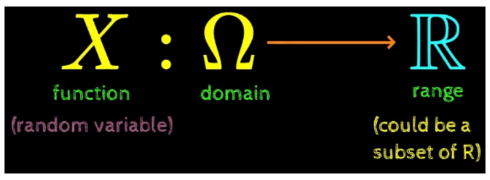


Fig.1 Random variable X is a function that maps the sample space to a range of real numbers.

Domain of the function (Ω): all the inputs the function can take

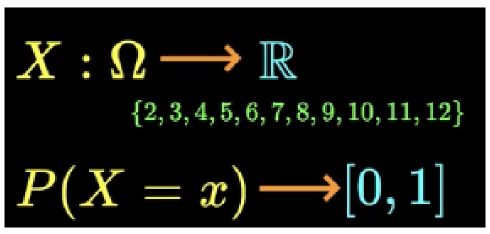
X : function that takes input from the sample space Ω and gives an output in R.

- **The function X is called a random variable.** It can take both discrete and continuous values.
- Multiple functions (random variables) can be mapped for a given sample space (domain).
- The function notation with brackets is not followed. Random variable are denoted using capital letters.
- **Discrete random variables** can take finite or countably infinite values. An example for countable infinite value is the number of coin tosses after which a head appears.
- **Continuous random variables** can take fractional values.

Probability Mass Function

def distribution - (t = 1:28)

- Using probablility mass function, the probabilities of the values that the random variable can take can be found.



X maps Ω to R, Function P maps the set X to the interval [0,1].

P is Probability Mass Function (PMF), a.k.a Probalility distribution / distribution.

Fig.2 The probability density function P gives the distribution of X

- The assignment of probabilities to all possible values that a discrete RV can take is called ***distribution*** of discrete random variable.
- The key idea is to think of the event coressponding to X = x, from this P(X = x) can be computed using the following:

$$p_X(x) = P(X = x) = P(\omega \in \Omega : X(\omega) = x)$$

- The following is an example of finding probability distribution for sum of outcomes while rolling two dies. Here the sum is the random variable and the die rolls form the domain.

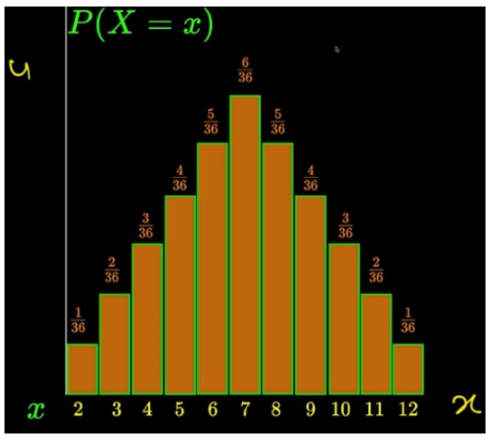


Fig.3 P(X = sum of outcomes of two die rolls)

Properties of PMF

def support of RV- (t=2:55)

1. The output of PMF should always be greater than or equal to 0.

$$p_X(x) \geq 0$$

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}) \geq 0$$

2. Sum of all the probability densities is equal to 1.

$\sum_{x \in \mathbb{R}_X} p_X(x) = 1$

Proof:

$\sum_{x \in \mathbb{R}_X} p_X(x) = \sum_{x \in \mathbb{R}_X} P(X = x)$

RHS is the sum of the probabilities of disjoint events which partition Ω

Sum of all the disjoint poprbabilities is 1.

- **Support** of a RV is the set of values that the random variable can take.

Discrete Distributions

The focus of the lecture is on PMF of discrete random variables.

> **Recap :** Distribution of a random variable (RV) is an assignment of probabilities to all possible values that a discrete RV can take.

Can PMF be specified compactly? (rather than using an elaborate table)

- Using the following coin toss example, it can be observed that listing down all the possible outcomes is not desirable. Thus, the succinct expression on the right is used the express the required PMF.

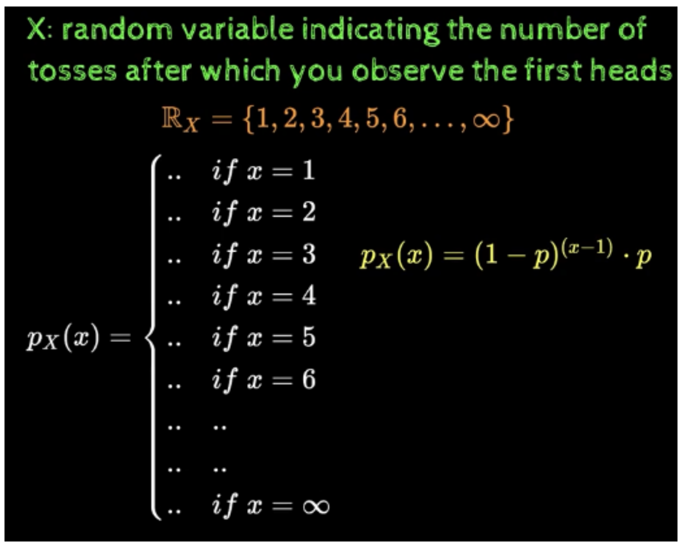


Fig.4 here p: probability of heads

- the function is **compact, easy to compute** and also **no enumeration is required.**
- In machine learning, the distribution is specified by some parameter. The parameters of the complex function is learnt from data.

Bernoulli Distribution

def Bernoulli trails - (t = 0:50)

- Experiments which have only two outcomes are called **Bernoulli trials.** One outcome can be considered as success(1) and the other a failure(1). The RV mapping the input to the binary outcome is called Bernoulli RV.

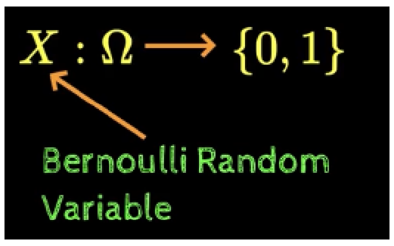


Fig.5 Bernoulli RV.

- The PMF of Bernoulli trials can be given as follows:

$$pX(x) = p^x(1-p)^{1-x}$$

where,

$$pX(1) = p$$
$$pX(0) = 1 - p$$

- **Is Bernoulli distribution a valid distribution? i.e.** $pX(x) \geq 0$

It is known that $0 < p < 1$ where p is the probability of success and $1 - p$ is the probability of failure. Thus, $pX(x) \geq 0$ is satisfied.

- Is $\sum_{x\epsilon(0,1)} pX(x) = 1$?

$$\sum_{x\epsilon(0,1)} pX(x) = pX(0) + pX(1)$$
$$(1-p) + p = 1$$

Binomial Distribution

- A Bernoulli trail is repeated n times such that the outcomes are **independent** (success/failure in one trial does not affect the outcome of other trials), **identical** (probability of success 'p' is each trial is the same).

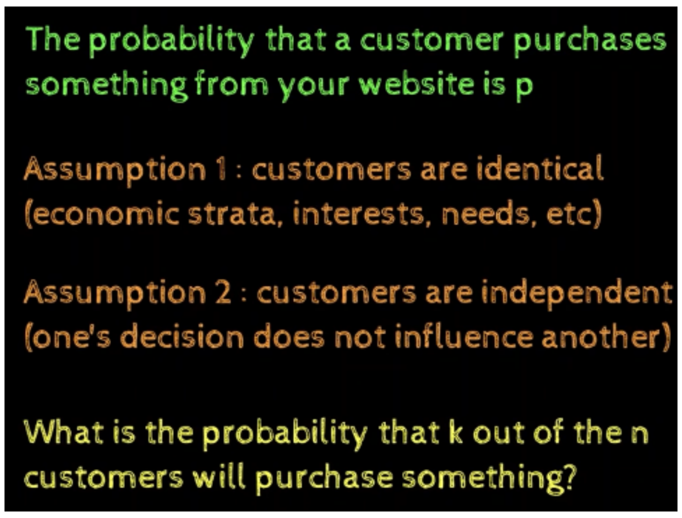


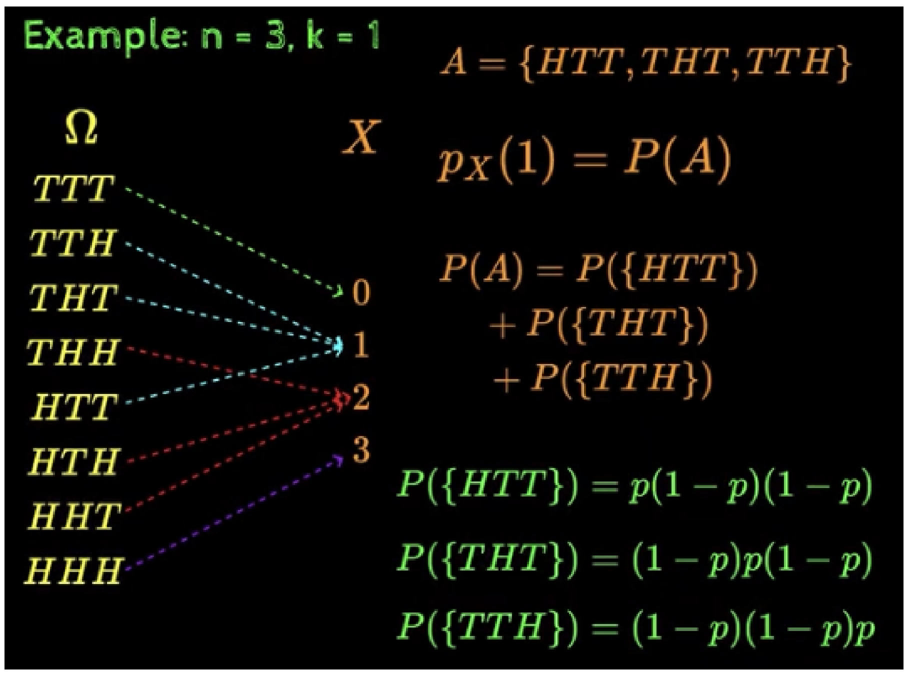
Fig.6 Example of a binomial distribution.

- The goal is to find the probability of **k** successes in **n** trials ($k\epsilon[0,n]$) in terms of n (number of trials) and p (probability of success).

Example (Binomial Distribution)

- **How many different outcomes can we have if a Bernoulli trial is repeated n times?**

2^n, where n is the number of times the experiment is repeated.



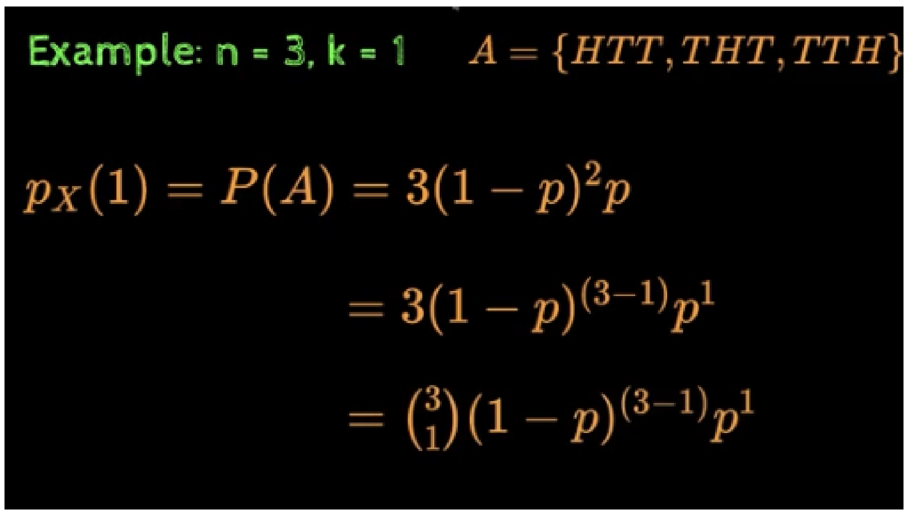
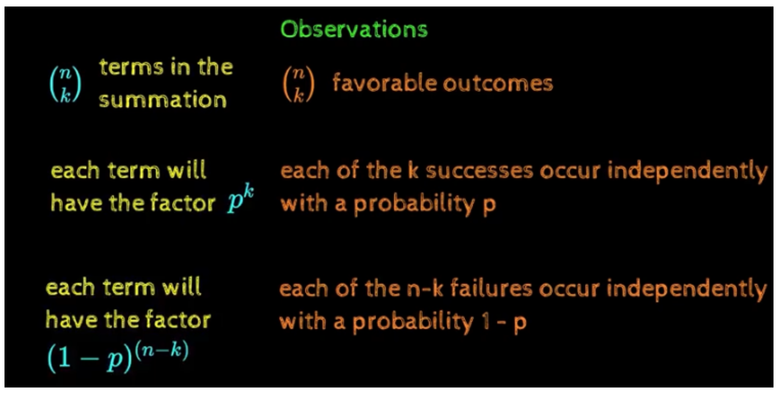


Fig.7 a,b Example of Binomial distribution and c. observations made



- Therefore, for given parameters p,n the PMF of binomial distribution is given by :

$$pX(k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

More Examples (Binomial Distribution)

- An example being infected by covid-19:

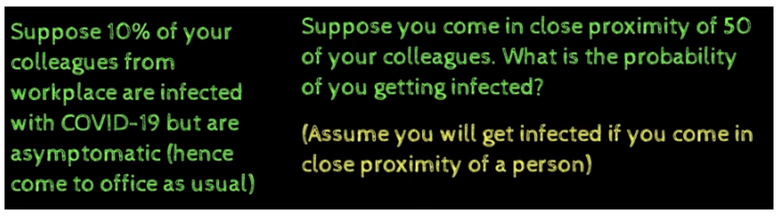


Fig.8a Example of Social distancing

Trial : Come in close proximity of a person.

p : 0.1 (probability of success/infection in a single trial)

n : 50 trials

Using the rule of subtraction,

$$
\begin{aligned}
P(getting\ infected) &= P(at\ least\ one\ success) \\
&= 1 - P(0\ successes) \\
&= 1 - p_X(0) \\
&= 1 - \binom{50}{0} p^0 (1-p)^{50} \\
&= 1 - 1 * 1 * 0.9^{50} = 0.9948
\end{aligned}
$$

Fig.8b Solution for 8a.

Change in p/n just requires the substitution to be changed in the RHS.

- The next example is solved to give a graphical intuition of the binomial distribution, using seaborn and scipy

```
# example code
import seaborn as sb
import numpy as np
from scipy.stats import binom

x = np.arange(0, 25) # the values RV can take
n = 25
p = 0.1
dist = binom(n, p)
ax = sb.barplot(x = x, y = dist.pmf(x))
```

Q : 10% students in a class use Linux. if 25 are selected at random then

a. what is the probability that exactly 3 use Linux?

sol : here, n = 25, p = 0.1, k = 3

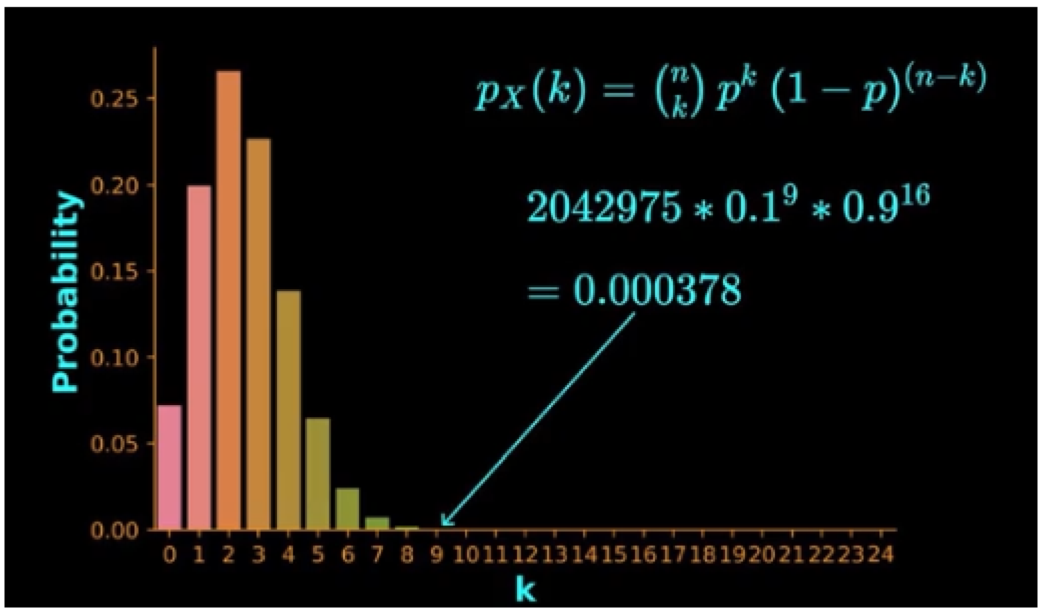


Fig.9a As a lower percentage of students use Linux, the graph is left skewed.

b. What is the probability that between 2-6 of them use Linux?

sol: here, n = 25, p = 0.1, k = (2,3,4,5,6)

c. How would the above probabilities change is instead of 10%, 90% or 50% of the students used Linux?

sol: here, n = 25, p = 0.9, p = 0.9 or 0.5

For p = 0.9, the plot would shift to the right and, for p = 0.5 the plot would shift to the center.

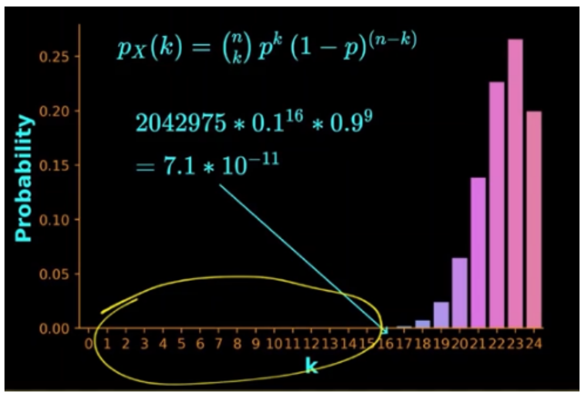


Fig.9b As higher percentage of students use the OS, the graph has a left tail.

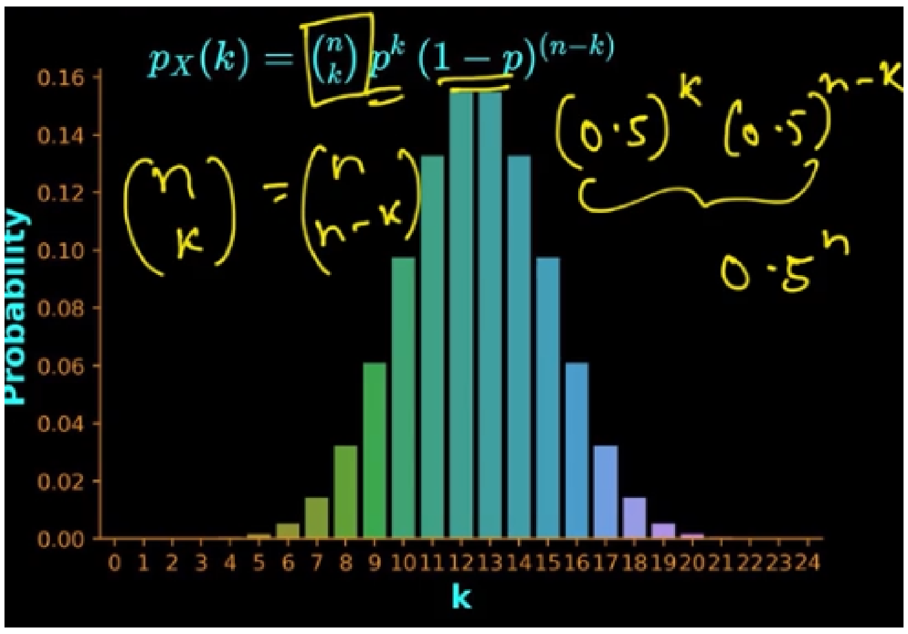


Fig.9c A symmetric distribution is observed.

Is Binomial Distribution a Valid Distribution?

- to show : $pX(x) \geq 0$

since p is assumed to always be positive, p^n and $(1-p)^{(n-k)}$will always be positive. Thus, $pX(x) \geq 0$ is always positive.

- Is the sum of all the values of probability that the RV can take equal to 1?

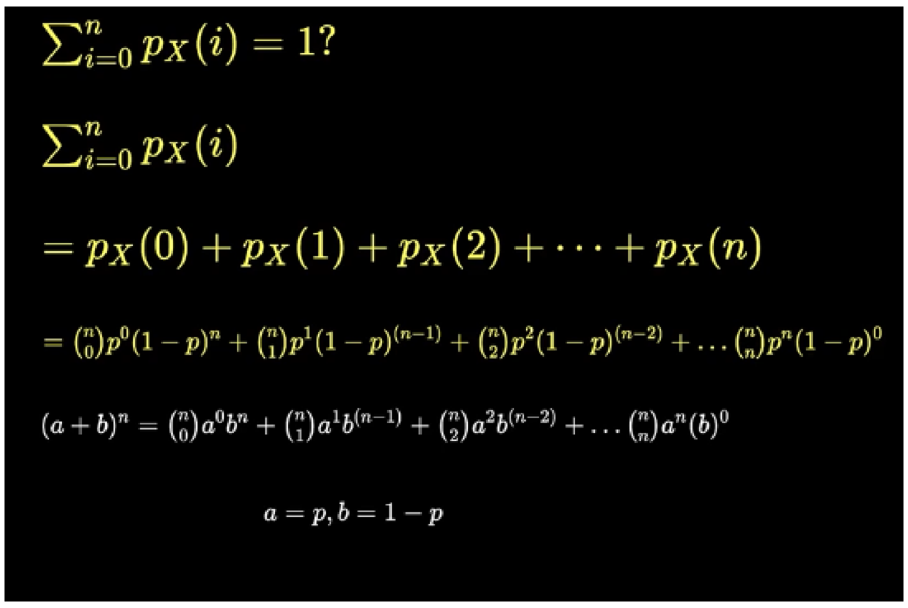


Fig.10 Proof for the second property.

Here, substituting the a and b values in the corresponding LHS we get 1.

- Bernoulli distribution is a special case of binomial distribution where, n = 1 and k can only take values {0,1}.

Binomial equation is given by

$$pX(k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

substituting, k = 0,1 we get the Bernoulli Distribution

$$pX(0) = \binom{1}{0} p^0 (1-p)^1 = 1-p$$
$$pX(1) = \binom{1}{1} p^1 (1-p)^0 = p$$

Geometric Distribution

- It involves doing a Bernoulli's trial infinite times, calculating the number if tosses until first success is seen. E.g. The coin is tossed infinite times (instead of n times), number of tosses until the first head is seen
- **motivation : Useful in any situation involving "waiting times".**

> To compute the chance that the first success will occur after k trials.

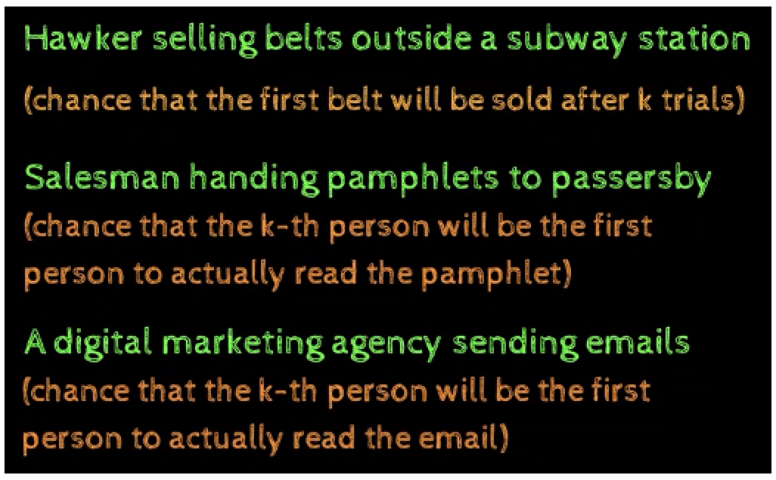


Fig.11 examples where Geometric distributions are used.

- Assumptions made are that the trials are **independent** and have an identical distribution.
- The PMF of success in the k^{th} trial, $pX(k)$ is given by

$$pX(k) = (1-p)^{k-1}p$$

- As k increases, the probability of (k-1) failures decreases. Irrespective of the values of p the geometric distribution always has a right tail.

Is Geometric Distribution a Valid Distribution?

- to show : $pX(x) \geq 0$

since p is assumed to always be positive, and $(1-p)^{(k-1)}$will always be positive. Thus, $pX(k) \geq 0$ is always positive.

- Is the sum of all the values of probability that the RV can take equal to 1?

$\sum_{k=1}^{\infty} pX(i) = 1?$

LHS = $(1-p)^0 p + (1-p)^1 p + (1-p)^2 p + ...$
LHS = $\sum_{k=0}^{\infty}(1-p)^k p$ (this is of the form $a, ar, ar^2, ar^3 ....$)

$a = p$ and $r = 1 - p < 1$

- An example of donor list:

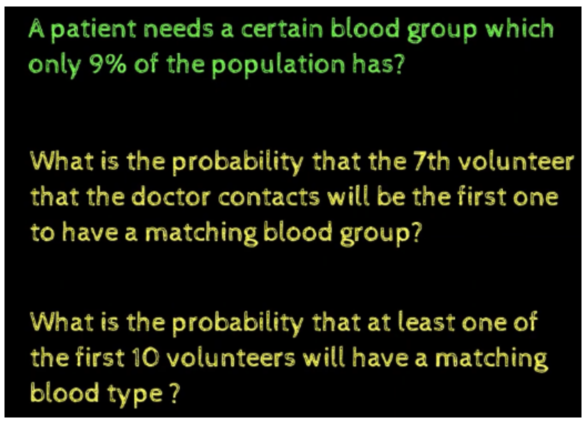


Fig.12 An example of geometric distribution.

Uniform Distribution

- These are experiments with equally likely outcomes
- If the distribution of x is uniform. PMF of x is given by

$$pX(x) = \begin{cases} \frac{1}{b-a+1} & \text{if } a \leq x \leq b \\ 0 & otherwise \end{cases}$$

- where (b - a + 1) gives the number of elements between a and b.

There are two special cases :

1. **a = 1 , b = n**

$$pX(x) = \begin{cases} \frac{1}{b-a+1} = \frac{1}{n} & \text{if } a \le x \le b \\ 0 & otherwise \end{cases}$$

i.e. each element of the sample space has equal probability of being selected.

2. **a = c, b = c**

$$pX(x) = \begin{cases} \frac{1}{b-a+1} = 1 & \text{if } a \le x \le b \\ 0 & otherwise \end{cases}$$

When RV is a constant, it's probability of being selected is 1.

Is uniform distribution valid?

- As long as $b \ge a$, $\frac{1}{b-a+1}$ is positive. thus, $pX(x) \ge 0$
- $\sum_{i=a}^{b} pX(i) = 1$?

$$LHS = \sum_{i=a}^{b} \frac{1}{b-a+1} = (b-a+1)\frac{1}{b-a+1} = 1$$

hence proved.

Expectation

def expectation - (t = 12:31)

- Expected value of the RV gain : an example of gambling.

Fig.13 Example of calculating the average gain.

- Expectation of RV X is given by

$$E[X] = \sum_{i=1}^{n} x_i * p_X(x_i)$$

- The quantities are weighted based on their probability of occurrence, thus, making it different from average. If all outcomes are equally likely, then the expectation is same as the average value.

Examples - Expectation

- Expectations quantifies the long term average gain for a given RV .
- Two examples of insurance were discussed, one to calculate the expectation for a given Bernoulli trial and the other was to calculate the premium given the expectation.

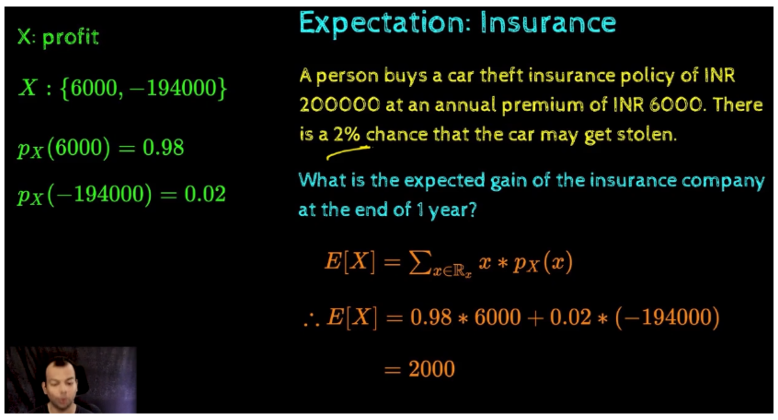


Fig.14a Example to calculate expectation for a given Bernoulli trial.

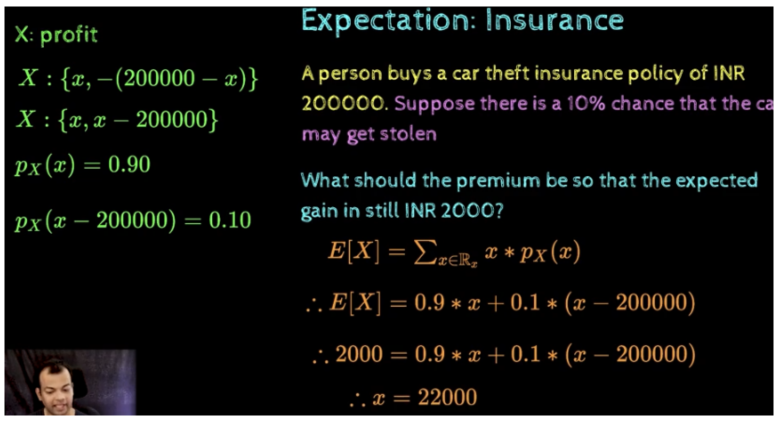


Fig. 14.b Example to calculate the premium for a given expectation.

Properties of Expectation

- Linearity of expectation

if $Y = aX + b$ then, $E[Y] = aE[X] + b$

- Given a set of random variables $X_1, X_2, ..., X_n$

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

- Expectation as mean of population

Every element in the population is equally likely to be selected, i.e. the RV has a uniform distribution. Then, the expected value is same as the mean.

- Expectation is also the centre of gravity for the PMF of a given RV

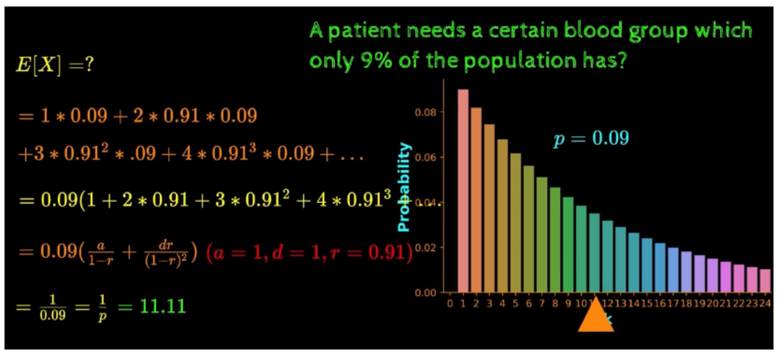
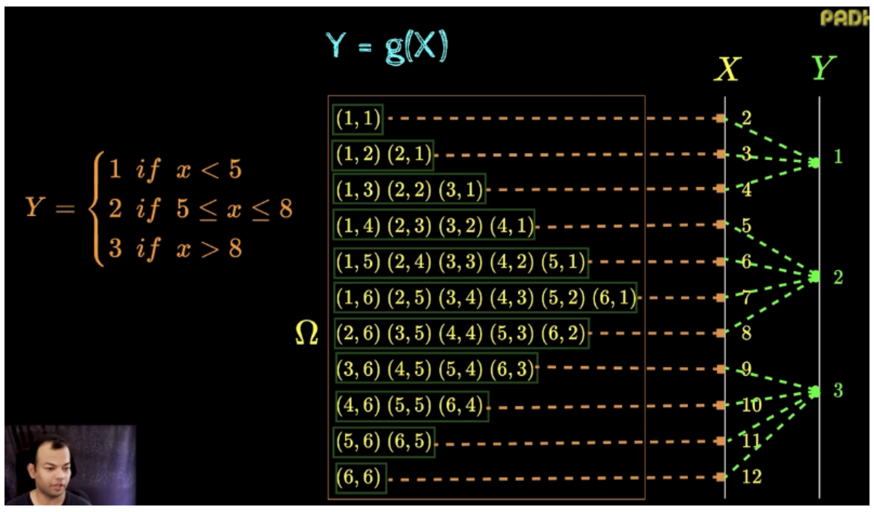


Fig.15 Example of expectation as the centre of gravity.

Functions of a Random Variable

- The value that function of a RV value takes will also be random
- Ex. $Y = g(X)$



$$Y = \begin{cases} 1 \text{ if } x < 5 \\ 2 \text{ if } 5 \le x \le 8 \\ 3 \text{ if } x > 8 \end{cases}$$

Fig.15a An example of function Y of a RV X.

- $E[Y]$ depends on the RV X. The PMF (distribution) of X is known.

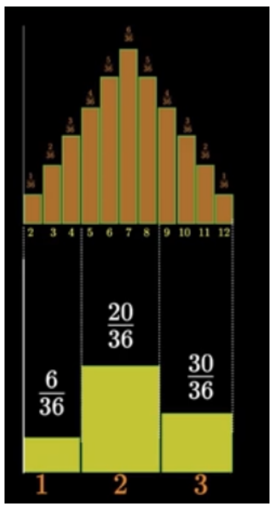


Fig.15b Equivalence of the distributions.

$$E[Y] = \sum_y g(x) * p_X(x)$$

here, $g(x)$ corresponds to the outcomes of RV Y and . The above equation is equivalent to $E[Y] = \sum_y y * p_Y(y)$.

Variance of a Random Variable

- Expectation summarises a RV but does not give information about its spread.

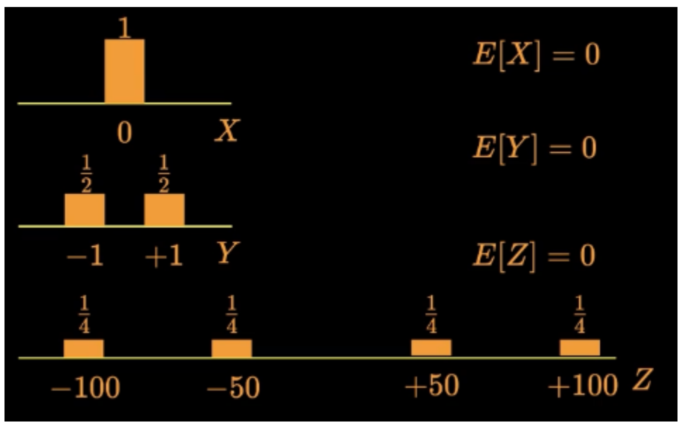


Fig.16 The 3 RV have the same expectation but the spreads are visibly different.

Recap: Variance is the average of the squared distance between all points and the mean.

$$Var(X) = E[(X - E(X))^2]$$

- here, 'X' the values of RV corresponds to the values of the population $x_i$and $E[X]$corresponds to the mean μ in the variance formula $\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$. The expected variance is the weighted sum of distance of all values of RV X from the expected value $E[X]$.
- The simplified formula of variance is

$$Var(X) = E[X^2] - (E[X])^2$$

where $g(X) = X^2$.

- Standard deviation of a RV $\sigma(X) = \sqrt{Var(X)}$

Properties of Variance

def independence of n-RV - (t = 6:57)

- $Var(aX+b) = a^2Var(X)$
- In general, **variance of the sum of RV** is not equal to the **sum of the variances of RV.** This is true when the RV are independent.

$$P(X=x|Y=y) = P(X=x)$$

then, X and Y are independent.

- If n random variables are independent, then

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

Summary

- Random variable is a function that maps the domain Ω to the a range $\mathbb{R}$, the scope of the RV.

$$X_{function} : \Omega_{domain} \rightarrow \mathbb{R}_{range}$$

- PMF is assigning probability values to each value a RV can take. PMF should satisfy the conditions $p_X(x) \geq 0$ and $\sum_{x \in \mathbb{R}_X} p_x(x) = 1$.
- The PMF of Bernoulli RV is given by, $p_X(x) = p^x(1-p)^{1-x}$.
- When a Bernoulli trial is repeated n times it is called binomial RV. Here the number of successes for n trials is calculated. The PMF of binomial RV is given by $p_X(k) = \binom{n}{k}p^k(1-p)^{(n-k)}$. The plot for the PMF can be left skewed, right skewed or symmetric, depending on the probability p of success.
- When an experiment is repeated infinite times, the probability of observing the first success at k^{th} trial is given by $p_X(k) = (1-p)^{(k-1)}p$. Irrespective of the probability of success p, the plot always has the same trend.
- When all the values of a RV are equally likely , the PMF follows a uniform distribution.
- Expectation is the weighted sum of the RV given by $E[X] = \sum_{i=1}^{n} x_i * p_X(x_i)$. It signifies the long term value of the RV.
- The expectation of a function of RV $Y = g(X)$, can be calculated in terms of the PMF (distribution) of the RV X.
- Variance of a RV is given by $Var(X) = E[X^2] - (E[X])^2$.
- If $Y = aX + b$ then, $E[Y] = aE[X] + b$ (both have a linear relationship) and $Var(Y) = a^2Var(X)$.
- Expectation is the center of gravity and variance captures the spread in the data.
- In the case of n RV, $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$ only when the RV are independent.

MCQ

1. A variable that can assume any value between two points is called:
    1. Discrete random variable
    2. Discrete sample space
    3. Random variable
    4. **Continuous random variable**
2. If X~N(μ,σ2) and a and b are real numbers, then mean of (aX+b) is
    1. a+b
    2. **aμ+b**
    3. aμ
    4. a+bμ

3. In a Binomial Distribution, if ‘n’ is the number of trials and ‘p’ is the probability of success, then the mean value is given by
    1. **np**
    2. n
    3. p
    4. np(1-p)
4. n a Binomial Distribution, if p, q and n are probability of success, failure and number of trials respectively then variance is given by
    1. np
    2. **npq**
    3. np^2q
    4. npq^2
5. If p = q, then P(X = x) for a binomial distribution is given by
    1. **nCx (0.5)n**
    2. nCn (0.5)n
    3. nCx p(n-x)
    4. nCn p(n-x)
6. Binomial distribution is
    1. not a probability distribution
    2. continuous distribution
    3. **discrete distribution**
    4. irregular distribution
7. A programmer has a 95% chance of finding a bug every time she compiles his code, and it takes her three hours to rewrite the code every time she discovers a bug. Find the probability that she will finish her program by the end of her workday. (Assume that a workday is 9 hours)
    1. 76%
    2. 44%
    3. 37%
    4. **28%**
8. What is variance of a geometric distribution having parameter p=0.80?
    1. **31%**
    2. 13%
    3. 67%
    4. 54%